

# UCSF

## UC San Francisco Previously Published Works

### Title

An ontology of randomized controlled trials for evidence-based practice: content specification and evaluation using the competency decomposition method

### Permalink

<https://escholarship.org/uc/item/0fk7d5k0>

### Journal

Journal of Biomedical Informatics, 37(2)

### ISSN

1532-0464

### Authors

Sim, Ida  
Olasov, Ben  
Carini, Simona

### Publication Date

2004-04-01

Peer reviewed

**An Ontology of Randomized Controlled Trials for Evidence-Based Practice:  
Content Specification and Evaluation Using the Competency Decomposition**

**Method**

Ida Sim MD, PhD, Ben Olasov MS, Simona Carini MA

Running Head: Clinical Trials Ontology Specification and Evaluation

From the Department of Medicine and the Program in Biological and Medical Informatics,  
University of California, San Francisco, CA.

Corresponding Author,     Ida Sim MD, PhD  
and Author for Reprints:   Assistant Professor, Department of Medicine, and  
                                  Associate Director for Medical Informatics, Program in Biological  
                                  and Medical Informatics  
                                  University of California, San Francisco  
                                  3333 California St., Suite 435 Q  
                                  San Francisco, CA 94143-1211  
                                  Phone: (415) 502-4519     FAX: (415) 514-0425  
                                  E-mail: [sim@medicine.ucsf.edu](mailto:sim@medicine.ucsf.edu)

Word Count: 5712

**Abstract**

Randomized controlled trials (RCTs) are one of the least biased sources of clinical research evidence, and are therefore a critical resource for the practice of evidence-based medicine. With over 10,000 new RCTs indexed in Medline each year, knowledge systems are needed to help clinicians translate evidence into practice. Common ontologies for RCTs and other domains would facilitate the development of these knowledge systems. However, no standard method exists for developing domain ontologies. In this paper, we describe a new systematic approach to specifying and evaluating the conceptual content of ontologies. In this method, called competency decomposition, the target task for an ontology is hierarchically decomposed into subtasks and methods, and the ontology content is specified by identifying the domain information required to complete each of the subtasks. We illustrate the use of this competency decomposition approach for the content specification and evaluation of an RCT ontology for evidence-based practice.

**Keywords:** ontology, clinical trials, knowledge representation, systematic review, electronic publishing, decision support systems, evidence-based medicine

Word count: 149

## I. INTRODUCTION

Ever since the notion of evidence-based practice was first articulated in 1992 (1), it has become increasingly accepted that medical care should be based as much as possible on the best available evidence from scientific research rather than on expert opinion or the physician's own personal experience (2). This ascendance of evidence over eminence is also reflected in the design of clinical decision support systems (CDSSs), the majority of which now base their recommendations on published evidence rather than on expert knowledge (3). That the medical literature is playing an increasingly central role in clinical practice has profound implications for how clinical research should be published: rather than publishing research findings in traditional prose articles that computers cannot understand (4), the scientific community should publish research findings directly into machine-understandable knowledge bases that CDSSs can use to help clinicians translate evidence into practice.

In the Trial Bank Publishing Project (5), we have been collaborating with *JAMA* and the *Annals of Internal Medicine (Annals)* to co-publish randomized controlled trial (RCT) reports as both prose articles and as entries into a machine-understandable "trial bank" called RCT Bank (6, 7). RCTs are one of the most valuable sources of evidence for clinical practice. With over 10,000 new RCTs indexed annually in Medline, an open-access knowledge base of the RCT literature is an important and necessary piece of infrastructure for CDSSs that aim to support evidence-based practice.

In developing RCT Bank to capture detailed information about the design, execution, and results of RCTs, the specification of an RCT ontology was critical. However, no standard method exists for developing domain ontologies (8). In this paper, we describe a new ontology specification and evaluation approach called competency decomposition and demonstrate its use

for designing our RCT ontology. We also describe the ontology's implementation, and report on an evaluation of its ability to capture RCTs from the literature.

## II. BACKGROUND

### A. From Evidence to Action

Because RCTs yield the least biased information on the effectiveness of interventions, RCTs are often considered the gold standard source of evidence for evidence-based practice. However, substantial gaps often exist between every day practice and "best practice" as defined by RCTs (9, 10). One reason for these gaps is that interpreting and applying RCT evidence is neither quick nor simple. The long road from evidence to action begins with interpreting and summarizing all the RCTs that are relevant to a particular clinical question (Figure 1). The canonical method for trial interpretation and summarization is *systematic reviewing* (11), which involves assessing the quality of a set of studies, exploring statistical heterogeneity, statistically combining the results if appropriate, and assessing the generalizability of the results to various patient populations (12). To enable computers to support evidence-based practice, therefore, RCT knowledge bases for evidence-based practice must first and foremost support the task of systematic reviewing. We designate as "trial banks" only those RCT knowledge bases whose *target task* is systematic reviewing. For such trial banks, their ontologies should be driven by the information requirements of systematic reviewing.

### B. Information Requirements of Systematic Reviewing

The trial information required for systematic reviewing has been partially enumerated in recommendations on what trial reports ought to contain (13-17). Almost all the recommendations

call for the reporting of details about trial design (e.g., participant blinding), execution (e.g., number of participant withdrawals in each intervention group), and results (e.g., denominators for all values). However, none of the recommendations enumerated the full set of required trial information because they were all implicitly bound by the assumption that trial reports are constrained in size. For example, the CONSORT trial reporting recommendation, accepted by over 80 journals worldwide (18), includes only 22 items out of the over 100 items they describe as needed for systematic reviewing (19). Despite these shortcomings, we drew heavily on the trial reporting recommendations as a starting point for our ontology specification.

### C. Other RCT Data Models

To further inform the development of our RCT ontology, we reviewed the data models of other trial databases and knowledge systems. The Cochrane Controlled Trials Register (20), although often referred to as a trial database, is more accurately a bibliographic database of trial report citations. Trial databases and registries such as ClinicalTrials.gov (21), *m*RCT (22), PDQ (23), and others (24) generally include only information sufficient for recruiting subjects or for determining the existence of a trial (e.g., entrance criteria, interventions, contact details), and thus are not “trial banks” in our definition. These databases are also mostly text databases, with XML structures but not formal data models. Few of the databases modeled trial protocols beyond text descriptions of the entrance criteria, and none modeled trial follow-up or outcome results. The data modeling from these databases was therefore of limited value for specification of our RCT ontology.

Of greater value was our examination of data models from trial execution (25, 26) and trial design (27) software. These data models ranged from entity-relationship and UML-specified

models to frame-based ontologies. While these models had useful representations of eligibility rules and time (28), which we borrowed for our ontology, they did not contain sufficient information about study methods (e.g., allocation concealment) as would be needed by meta-analysts.

## III. ONTOLOGY SPECIFICATION

### A. Method

We specified the contents of and evaluated our trial-bank ontology using a method we call *competency decomposition*. In this method, a top-level *target task* is decomposed into *subtasks* and the *methods* by which each subtask is to be accomplished. Then, the necessary and sufficient trial information items are defined for completing each subtask using the specified method(s). This decomposition process is recursively iterated across all the subtasks, thus generating a task hierarchy whose leaves are the required information items for the lowest level subtasks. If an ontology has all the requisite information items for a task, it is said to be competent for that task. For example, quantitative synthesis is a target task of trial banks (Table 3). The Mantel–Haenszel formula is one method for statistically pooling observed odds ratios into a summary estimate. The Mantel–Haenszel formula requires the odds ratios for each of the trials to be combined (data requirement II.A.2.a), which implies a requirement for the complete 2 X 2 contingency table for each outcome of each trial that is to be pooled (data requirement II.A.1.a). This competency decomposition leads to two implications for trial banks that intend to support Mantel–Haenszel meta-analysis of odds ratios. First, their data schemas must model a complete 2 X 2 contingency table for each outcome variable. Second, trial authors must report complete 2 X 2 tables for each

outcome in their trials. Using this framework, the data requirements for other meta-analytic methods can be similarly specified.

Our competency-decomposition method adapts and combines Chandrasekeran's Generic Tasks (29) and Gruninger's competency-questions (30) approaches to knowledge system design. The key contribution of the Generic Tasks idea was to represent knowledge based on how that knowledge was to be used in a knowledge system. Generic Tasks (e.g., classification, data retrieval, plan selection and refinement) were defined that could be implemented as problem-solving "building blocks" for intelligent systems. Tasks could be further decomposed into subtasks, and the methods by which the knowledge system will achieve them, to drive more specific problem-solving modules. We adapted this the task decomposition approach to specify the content of domain ontologies rather than problem-solving modules. For the competency-questions approach, the key contribution was a formal approach to specifying and evaluating the "competencies" of a knowledge system, or what a knowledge system claims to be able to do. We improved upon this approach by providing a hierarchical task-based approach to defining the competency questions.

While our adaptations of the Generic Task and competency-questions approaches are not large, the resulting benefits of our competency decomposition approach are significant. First, a task hierarchy is understandable by both knowledge-modeling and domain experts, and can thus be revised easily in collaboration. Second, it is clear which information items are required for which tasks, and also which tasks each information item is required for. Finally, the competency decomposition can be used to specify, evaluate, or document an ontology. An ontology's content specification is simply the information requirements for the tasks and subtasks that the knowledge base is to support. An ontology can be evaluated by comparing its content to the

competency decomposition: the ontology is "competent" for all tasks for which it contains all the information requirements. An ontology can be documented by listing the tasks for which an ontology is competent. Such documentation can promote the sharing and reuse of ontologies, and can facilitate the definition of common ontologies. These benefits stand in contrast to other methods for specifying ontology content. For example, neither deriving an ontology's content from ad hoc cogitation nor from natural language processing of a corpus leads to clear statements about an ontology's purpose or what tasks it can support. Such clear and task-based statements of ontological "competencies," as can be provided by our method, are valuable for using shared ontologies to improve knowledge systems.

Our competency decomposition method can be applied to any domain. We now illustrate its application to the modeling of RCTs.

## **B. The Trial Bank Competency Decomposition**

Using systematic reviewing as the target task for our competency decomposition, we identified 5 top-level tasks, 23 subtasks, and 74 sub-subtasks based on the literature and on personal experience with systematic reviewing (<http://retbank.ucsf.edu/tasks/tasks.html>). The five top-level tasks are: (1) trial retrieval, (2) judgment of each trial's internal and (3) external validity, (4) quantitative synthesis of trial results, and (5) interpretation of the trial within the epidemiologic and socioeconomic context of the clinical problem under investigation (Table 1).

*Task 1: Trial retrieval.* A systematic review starts with a comprehensive search for all relevant trials (31, 32). Queries for systematic reviews often search on trial concepts such as inclusion and exclusion criteria, subject characteristics, interventions, and outcomes. An example

query is “all RCTs on aspirin use in women over age 65 that included more than 6 months of follow-up, in which the primary outcome was total mortality and the follow-up was over 80%.” To support trial retrieval, our competency decomposition identifies the trial concepts that should be modeled in a trial-bank ontology, but not particular instantiations of those concepts. For example, while the competency decomposition specifies Intervention as a needed concept, it does not dictate which particular interventions (e.g., Aspirin) need to be modeled. For RCT Bank, we use the UMLS (Unified Medical Language System (33)) for coding the clinical semantics of attribute values. We chose the UMLS preferred terms as our controlled vocabulary because of the UMLS’ scope, availability, and low cost. In our competency decomposition, no information requirements are listed under this task decomposition because the retrieval tasks all overlap with other top-level tasks.

*Task 2: Judging internal validity.* A trial’s internal validity is the extent to which its findings are an unbiased estimate of the effect under study. Although there are no standard methods for judging internal validity, methodologists generally agree on the trial information they need for making these judgments (34). For example, because low compliance rates lead to underestimates of an intervention’s true effect, all methodologists would agree that information on compliance is needed for judging internal validity even though they may disagree on how to adjust for low compliance (35, 36). We identified and decomposed 11 subtasks and 40 sub-subtasks for judging internal validity independent of any particular method, but with explanations for how the trial attributes may affect internal validity. Overall, 112 information items were required for this task.

*Task 3: Judging external validity.* A trial’s external validity is the extent to which its findings are generalizable beyond the immediate circumstances of the trial. For example, to determine if a trial is relevant to a particular patient, one needs to know where the trial was performed and its inclusion and exclusion criteria. As with the judgment of internal validity, there are few standard methods for judging external validity, so the 4 subtasks and 14 sub-subtasks were also decomposed independent of any particular method for evaluating external validity. Overall, 29 information items were required for this task.

*Task 4: Quantitative synthesis.* If a set of trials is sufficiently similar in their methodology and clinical features, it may be appropriate to pool their results statistically in order to achieve greater power for detecting an effect. The subtasks of quantitative meta-analysis are to determine the design, clinical, and statistical heterogeneity of the trials, to calculate a summary statistic for each trial (e.g., an odds ratio), and to combine these numbers using a meta-analytic method if appropriate. Deciding whether trials are sufficiently similar to warrant meta-analysis is something of a black art, and the decomposition here is both method independent and sketchy. In contrast, there are many meta-analytic methods. We decomposed the tasks for the fixed effects, the random effects, and the meta-regression methods, and identified 4 unique required information items.

*Task 5: Contextual interpretation.* RCTs should be analyzed within their scientific, epidemiologic, socioeconomic, and ethical context. For our competency decomposition, we identified 3 subtasks and 12 sub-subtasks independent of any method for interpreting trial context. Eight of the sub-subtasks specify trial-related information needed for contextual

interpretation, information such as Institutional Review Board approval, informed consent procedures, potential competing interests of investigators, or whether findings were retracted for fraud or error. The other 4 sub-subtasks specify contextual information such as the epidemiology and treatment costs of the disease. Overall, 30 information items were required for this task.

This Trial Bank Competency Decomposition sets the content standards for the ontology of any knowledge system that purports to support systematic reviewing. Any particular ontology may fall short of this standard due to modeling implementation problems or explicit system design choices.

### **C. Evaluation of Competency Decomposition**

Our competency decomposition specified 147 unique items of trial information (referred to as the C-D Items) that should be modeled in a trial bank ontology. These C-D items included all the items mentioned in the trial reporting recommendations we reviewed (13-17). Furthermore, as we described elsewhere (6), the C-D Items included 96% of the information items needed by 18 published trial critiquing instruments (37). Trial critiquing instruments are an external check on what trial information is needed for judging a trial's internal and external validity, although these instruments have limited scope because they were designed for use on trial reports of limited length. The C-D Item's almost complete coverage of trial critiquing items supports our claim that the C-D Items includes all the information needs of systematic reviewing. The 4% of trial critiquing items that were not in the C-D Items were mostly items related to trial reports (e.g., study title, abstract) rather than to trials themselves. Conversely, 35% of the C-D Items, especially those relating to judging external validity, were not used by any of the 18 trial critiquing instruments, reflecting the limited scope of these instruments. In summary, we

validated the C-D Items against the state-of-the-art trial reporting and critiquing literature, and demonstrated that the C-D Items are a more complete enumeration of the information needs of systematic reviewing than is available in the literature.

## **IV. RCT SCHEMA**

We implemented the C-D Items content specification as an Ocelot frame-based<sup>1</sup> ontology (38) called RCT Schema.<sup>2</sup> The RCT Schema class hierarchy is 7 levels deep with 188 frames and 601 unique slots. There are an average of 9.8 slots per frame, and thirteen frames (7%) have multiple parents. 187 of the 601 slots (31%) take other frames as values.

We developed RCT Schema iteratively using a middle-out ontology modeling approach (39), in which we constructed our task hierarchy from the "middle tasks out" rather than from the top-level tasks down or from the lowest-level tasks up. More than 147 frames were needed to represent the 147 unique information items because some of the more complex information items (e.g., description of the outcome) required several frames to describe (e.g., rate outcomes need a denominator, cost outcomes need a discount rate).

### **A. Class Hierarchy**

RCT Schema captures details about a trial's administration, design, execution, and results (Table 2). The class hierarchy is rooted in the class TRIAL. A trial consists of one or more experiments conducted on a group of subjects selected according to predefined criteria. One trial may give rise to several studies: a study on the trial's main hypotheses, and potentially many

<sup>1</sup> Frames are equivalent to classes in object-oriented systems, and slots are equivalent to attributes.

<sup>2</sup> An ontology or information model differs from a database schema in that an ontology is primarily a domain model for knowledge representation purposes, whereas a database schema is primarily a technical artifact for the efficient storage of instances in a particular database.

secondary and follow-up studies. Each study may give rise to many publications. This conceptualization of the trial as the unit of analysis differs markedly from databases such as the Cochrane Controlled Trials Register, in which trial reports rather than the trials themselves are the unit of analysis. The schema and its documentation can be viewed at <http://rctbank.ucsf.edu/ontology/outline/index.htm>.

### *1. Trial*

The TRIAL class has slots for frames that describe the information unique to the entire trial, such as the trial name and registry numbers. For each TRIAL, the main study is described by a MAIN-STUDY class, while secondary or companion studies such as follow-up or subgroup studies are described by one or more SECONDARY-STUDY classes. These MAIN-STUDY and SECONDARY-STUDY classes themselves contain pointers to other classes that describe each study's administrative details, clinical and scientific background, conclusions, and publication details. Examples of some administrative details that are captured include investigator names and project contributions; conflicts of interest; the role of funders in data analysis and reporting; informed consent procedures; and locations, and characteristics of planned, withdrawn, and final study sites (Table 2). Separate slots hold information about the clinical background to the study, relevant prior evidence, and what this study adds to the literature. The title and citations of study publications are also stored, as well as other publications that refer to this study (e.g., editorials, letters to the editor, retractions).

### *2. Protocol*

Each MAIN-STUDY or SECONDARY-STUDY of the trial is further described by an INTENDED-PROTOCOL and/or an EXECUTED-PROTOCOL class. These classes differ only in that EXECUTED-PROTOCOL contains pointers to the actual results of the trial. For clarity, we restrict our description in this paper to the EXECUTED-PROTOCOL class. This class organizes information about the study design, study groups and subgroups, inclusion and exclusion criteria, interventions and co-interventions, study outcomes and side effects, subject recruitment and enrollment, treatment assignment and randomization, and follow-up. Where appropriate, these slots have unrestricted multiple cardinality so that, for example, studies with multiple arms and an arbitrary number of outcomes can be accommodated.

### *3. Design*

The statistical design of the MAIN-STUDY is captured in the class SAMPLE-SIZE-CALCULATION, which contains slots for the study's power and sample size calculations. The power calculation is indexed to one of the study's OUTCOME<sup>3</sup> instances. Randomization and assignment are described with slots on blocking, stratification, methods (fixed or adaptive), and assignment ratios. Allocation concealment methods can also be described.

### *4. Populations*

RCT Schema models groups of subjects using the class POPULATION, which has slots for the population name, size, age and gender distribution, and the UMLS term that best describes the population. First-level subclasses of POPULATION are RECRUITED-POPULATION, EXCLUDED-POPULATION and ANALYZED-POPULATION.

---

<sup>3</sup> A common alternative term for study 'outcome' is 'endpoint.'



Subclasses of RECRUITED-POPULATION include classes for subjects who were screened, eligible, enrolled, and randomized. Subclasses of EXCLUDED-POPULATION include classes for subjects who were screened but not eligible, eligible but not enrolled, enrolled but not randomized, and randomized but excluded from intention-to-treat analysis. These classes have, in addition to the POPULATION slots, a slot that takes multiple instances of the class REASON, which describes each reason for subject exclusion and the number of subjects excluded for that reason.

Subclasses of ANALYZED-POPULATION include classes for subjects assigned to each intervention (STUDY-ARM-POPULATION), subgroups for analysis, and subjects who crossed over off protocol from their assigned intervention to one of the other interventions. These ANALYZED-POPULATION subclasses include a slot for the INTERVENTION-ARM (see Interventions below) to which the subjects were assigned.

#### *5. Entrance Criteria*

Standard age and gender criteria are captured in the AGE-GENDER-RULE class, while ethnicity and language criteria are captured in the ETHNICITY-LANGUAGE-RULE class. The ethnicity designations are as defined by the United States Office of Management and Budget, and include a field for whether ethnicity was self-identified.

Entrance criteria other than the above are modeled as CLINICAL-RULES. The CLINICAL-RULE class has slots for the number of subjects satisfying and not satisfying this rule, a UMLS term describing the rule, and pointers to a combination of RECURSIVE-RULES and BASE-RULES. For example, the rule that subjects must actively be smoking either a pipe or > 2 packs of cigarettes a day would be captured as a RECURSIVE-RULE of one BASE-RULE (“Is an

active smoker”) connected with AND to another RECURSIVE-RULE, this one consisting of two BASE-RULES (“smokes > 2 packs of cigarettes a day” and “smokes a pipe”) connected by OR (Figure 2).

With this modeling approach, RCT Schema can capture any arbitrary nesting of rules. However, RCT Schema currently captures BASE-RULES as text strings only, which limits the machine-understandability of the overall rule. Also, the modeling of temporal aspects of BASE-RULES (e.g., “chest pain within the last 6 hours”) is preliminary.

#### *6. Interventions*

RCT Schema can accommodate multi-armed trials. The INTERVENTION-ARM class has slots for a name, a UMLS term describing that arm, and the one or more INTERVENTIONS that are administered to subjects in that arm. Each INTERVENTION is further described with classes appropriate for that type of intervention: drug, procedure, device, or other. For example, the class DRUG describes the drug generic name, manufacturer, trade name, and adjustments. A separate DRUG-STEP class describes the drug dosage, route of administration, frequency, duration, titration goal, and monitoring, so that RCT Schema can model complicated drug regimens such as loading and maintenance doses. The PROCEDURE class allows description of the personnel conducting the procedure, as well as the setting (e.g., ambulatory surgical center or operating room).

Each trial can have several EXPERIMENTAL-ARMS, but only one COMPARISON-ARM. Comparison interventions can be any of the classes allowed for EXPERIMENTAL-ARM (i.e., DRUG, PROCEDURE, DEVICE, OTHER-INTERVENTION), or could be PLACEBO, USUAL CARE, or NO-TREATMENT. The COMPARISON-ARM class also has a slot for justifying

why this control is clinically and statistically appropriate. Each arm is the assigned INTERVENTION-ARM for its corresponding STUDY-ARM-POPULATION.

### 7. Outcomes

Each study can have multiple PRIMARY-OUTCOMES, SECONDARY-OUTCOMES, BASELINE-CHARACTERISTICS, and SIDE-EFFECTS. Each outcome is modeled as a continuous, categorical, proportionate, rate, cost, life-year, or scored-instrument outcome. Depending on the outcome type, other relevant information can be collected, such the discount rate and base year for cost outcomes. An OUTCOMES-ASSESSMENT class captures information on how the outcome was assessed (e.g., use of computed tomography of the head with contrast to assess stroke), the reproducibility and validity of the assessment method, who performed the assessment (e.g., CT scan read by one board-certified neuro-radiologist), and whether the assessor was blinded to the subject's assigned intervention and/or to ongoing results of the study.

Each outcome has information about its univariate, multivariate, and/or survival analyses as appropriate, linked to the timepoints at which the outcome was assessed. For example, the STATISTICAL-ANALYSIS-AND-RESULTS instance of an outcome describes the univariate tests and has pointers to results at each timepoint of assessment (e.g., 6 weeks and 6 months). A REGRESSION-ANALYSIS-AND-RESULTS instance for that same outcome may describe a multivariate analysis only for the 6-month timepoint. This modeling allows different timepoints of assessment for different outcomes in a trial, as well as providing full flexibility for representing the statistical analyses undertaken. For each trial, one of the OUTCOME instances is also the primary statistical outcome, the outcome upon which the trial was powered.

### 8. Outcomes Follow-up and Intervention Compliance

Although widely used to describe follow-up, the terms “withdrawal” and “dropout” mean different things to different people. RCT Schema dispenses with these terms and models follow-up as the number of subjects who had a particular outcome assessed at a particular timepoint. The literature generally reports follow-up only for the primary outcome at the longest timepoint, but in fact, the follow-up can differ across outcomes and timepoints. In RCT Schema therefore, each timepoint for each outcome has a slot for the follow-up number for each STUDY-ARM-POPULATION (and for each SUBGROUP-POPULATION if applicable). Furthermore, there is also a slot for the number of subjects who had their outcomes assessed *while being on* their assigned intervention (the on-intervention follow-up). The numbers of subjects who remained on their assigned intervention at each timepoint is also captured in RCT Schema. All together, this follow-up and compliance information fully specifies the proportion of compliant and non-compliant subjects who had each outcome assessed at each timepoint. This data allows for unambiguous reporting of both intention-to-treat and efficacy analysis results.

### 9. Results

Results are captured in a way that is specific to the outcome type (e.g., mean, standard deviation, median, and range are allowable for continuous outcomes). For intention-to-treat analysis, the denominators should be the follow-up numbers for each STUDY-ARM-POPULATION for each outcome at each timepoint. If these numbers are not available, then the denominator is assumed to be the number of subjects in each STUDY-ARM-POPULATION. For efficacy analysis, the denominators should be the on-intervention follow-up numbers, as

described above. Summary statistics (e.g., odds ratios, p-values, confidence intervals) are stored for each outcome, STUDY-ARM-POPULATION, and timepoint combination, for each univariate, multivariate, and/or survival analysis. Subgroup results are modeled exactly as above using SUBGROUP-POPULATION instead of STUDY-ARM-POPULATION. Patient-level data can be captured by defining instances of SUBJECT as a STUDY-ARM-POPULATION of size 1.

Other results captured by RCT Schema include the degree to which subjects were able to guess their assigned intervention (blinding efficacy), as well as the trial's conclusions, discussion, limitations, and clinical implications.

#### **B. Modeling of Clinical Content**

The ontology content specification includes only concepts that are common to all RCTs, and does not impose any requirements or standards on the representation of the *clinical* aspects of RCTs. For disease-specific trial banks, clinical concepts could be modeled to a finer granularity to capture standardization within that clinical domain. For example, the National Cancer Institute has specified Common Data Elements (CDEs) for nine types of cancers (40). An example of a CDE for breast cancer research is Menopausal Status, which defines various patient characteristics that determine menopausal status (e.g., hormone levels). To customize RCT Schema for breast cancer trials, Menopausal Status could be modeled as a subclass of BASELINE-CHARACTERISTIC, so that instances of STUDY-ARM-POPULATION or SUBJECT would have standardized instantiations of the characteristics determining menopausal status. Trial banks based on such an extended Breast Cancer RCT Schema would then share among them the semantics of menopausal status while also sharing with all other trial banks the

semantics of generic RCT concepts. This granular approach to clinical standardization allows trial banks to exploit common trial and disease semantics when appropriate without being burdened by those semantics when it is not appropriate.

To facilitate cross-comparisons among trial-bank entries, it is desirable that the clinical terms come from a standardized clinical vocabulary but the competency decomposition does not specify which vocabulary to use. The richer the clinical vocabulary used, the richer will be the clinical reasoning possible over the trial bank. We currently use UMLS to code the clinical content of eligibility rules, outcomes, and interventions as discussed above, but will be switching to SNOMED CT as it becomes more available.

#### **V. USE OF RCT SCHEMA**

We have used RCT Schema to capture 11 trials completely and parts of an additional 13 trials. RCT Bank entries were entered using the Bank-a-Trial secure website, and can be browsed on the web at RCT Presenter (see <http://rctbank.ucsf.edu/> for all software).

#### **A. Competency of RCT Schema**

RCT Schema was able to capture 144 of the 147 C-D Items (98%) from real trials, representing the necessary and sufficient information for 69 of the 74 (93%) lowest-level subtasks of the competency decomposition. The items that RCT Schema could not capture concerned only trials with run-in or washout periods and were not critical to the major tasks of systematic reviewing. This evaluation thus shows that RCT Schema is highly competent for its target task of systematic reviewing. Furthermore, it illustrates how a competency decomposition can be used to evaluate which tasks an ontology is competent for.

## B. Characteristics of Captured Trials

Nine of the fully captured trials are part of the Trial Bank Publishing Project and were recently published in *JAMA* or the *Annals*. These trial-bank entries averaged 352 KB in size with 834 instance frames, 685 number values, and 1039 string values. The other 2 fully captured trials included information from internal trial execution records and are thus less representative of the information typically available about a trial.

The captured trials were diverse in their clinical domains, interventions, outcomes, and results (Table 3). The interventions included multi-step drug regimens, invasive procedures, and counseling. Some of the trials had multiple interventions per treatment arm. The outcome types included dichotomous, continuous, categorical, and scored-instrument (e.g., Wechsler Adult Intelligence Scale) outcomes. Univariate, multivariate, survival, and regression analyses were captured, as well as both intention-to-treat and efficacy analysis results. Examples of other trial attributes that were captured and that are of special concern to systematic reviewers included participant dropout at any stage of recruitment and follow-up (e.g., after randomization but before any intervention was started), early stoppage, protocol changes during trial execution, and participants' guesses about which intervention they received. We were not able to test a few parts of RCT Schema (e.g., reliability and validity of study questionnaires) because that information was not reported in any of our co-published trials.

As we captured more trials into RCT Bank, we had to refine RCT Schema to accommodate idiosyncratic deviations from traditional trial design. For example, sample sizes are usually calculated in reference to a primary outcome. In one trial, however, the sample size was calculated for an outcome that was listed as a secondary outcome in the final trial report. We

therefore relaxed our modeling to allow sample size calculations to be calculated for any outcome, not just the primary outcome.

In several cases, we encountered idiosyncrasies that we could not capture gracefully. For example, in one study, subjects who failed to meet inclusion criteria were nevertheless included in an idiosyncratically defined “efficacy” analysis. Following best practices for trial design and reporting, these subjects should have been included in an intention-to-treat analysis or been excluded post-randomization. Because these “efficacy” results violated basic tenets about our results modeling, we chose not to enter these results into RCT Bank. The judgment of which “idiosyncrasies” should be reflected in the modeling of RCT Schema depends on a detailed understanding of RCT methods and analysis. We are working with an international group of editors and biostatisticians (CONSORT Plus) to adjudicate these decisions.

## C. Limitations

RCT Schema can capture essentially all the information in detailed reports of typical interventional trials. Some rarely reported items that RCT Schema cannot yet capture include nested subgroups, and follow-up rates and intervention compliance for individual subgroups. RCT Schema also cannot yet capture crossover or cluster-randomized trials, or trials with run-in and washout periods, but these trial designs are relatively uncommon. The modeling for capturing follow-up studies and participant-level data is complete but has not yet been fully tested.

By using UMLS as our standardized clinical vocabulary, RCT Schema shares UMLS's well-known problems with capturing clinical concepts accurately (41-44). For example, one trial's intervention was percutaneous coronary intervention, which was defined as either percutaneous

coronary angioplasty or a stent. However, percutaneous coronary intervention is not a UMLS term. To properly express this concept using existing UMLS terms for angioplasty and stent, a term-composition grammar (45) would be needed. As we switch our clinical vocabulary from UMLS to SNOMED CT, we will be exploring how we can better capture clinical concepts using SNOMED CT's compositional features for generating post-coordinated terms.

One other limitation to the current RCT Schema is that most of the eligibility rules we capture are not fully machine-understandable. We do capture age, sex, ethnicity, and language criteria in highly coded fashion but for general eligibility criteria, we are currently capturing base clauses only in free-form text. The generic modeling of eligibility rules has been a long-standing challenge (28, 46), but it is critical that RCT Schema support automated eligibility determination so that decision support systems for evidence-based practice can automatically select the most relevant trials in RCT Bank to apply to particular patients. We are currently exploring generic modeling approaches to capturing the complex temporal and comparative relationships in eligibility rules in computable form.

## VII. DISCUSSION

We have used a new ontology content specification method called competency decomposition to specify and evaluate RCT Schema, an ontology of RCTs for trial reporting and analysis. We showed that RCT Schema is competent for the vast majority of systematic reviewing subtasks and that it has performed well in capturing a range of RCTs. The benefits of our competency decomposition approach include the clear specification of a target task to ground ontological modeling, the tight coupling of domain modeling to the information needs of the

target task, and the ability to use the competency decomposition as a “yardstick” to measure any ontology's competence for the target task.

To our knowledge, RCT Schema is the most complete trial ontology for trial interpretation and application to clinical care. It shares some content with other RCT ontologies that are targeted towards the trial performance parts of the clinical trial lifecycle (Figure 1) (e.g., trial design (27) or protocol management and execution (25, 26)). Ideally, systems from across the clinical trial lifecycle would all interoperate, as envisioned by the National Electronics Clinical Trials and Research (NECTAR) network (47). Such interoperation would be facilitated by a common RCT ontology.

The competency decomposition approach could be used to define commonalities between RCT Schema and other RCT ontologies. For example, other major target tasks for clinical trial systems (e.g., protocol eligibility determination, participant tracking, Good Clinical Practice compliance) could be defined and decomposed, and the information requirements for all high-level tasks compared and reduced to a common set. The process of explicitly defining target tasks and their information requirements may help clarify the interoperation needs of various clinical trial systems. The process may also serve to document and justify any eventual common information model. We are working with the HL-7 protocol representation group (48) to blend elements of the competency decomposition approach into their use-case driven information modeling efforts. This general approach should also be applicable to the modeling of other domains besides clinical trials.

## Acknowledgments

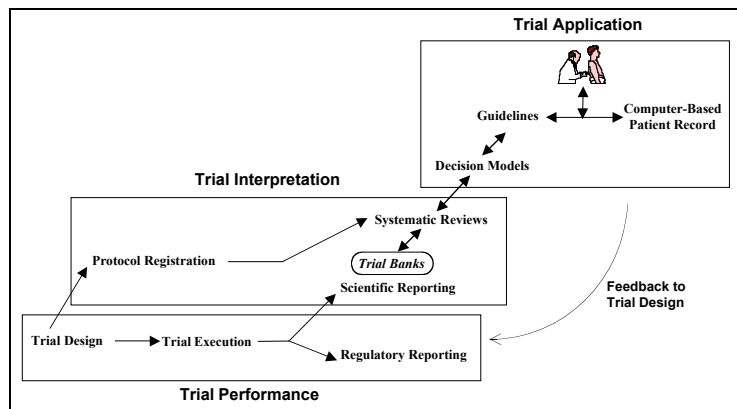
**This work was supported by grant LM06780 from the National Library of Medicine, and Dr. Sim's United States Presidential Early Career Award for Scientists and Engineers.**

## References

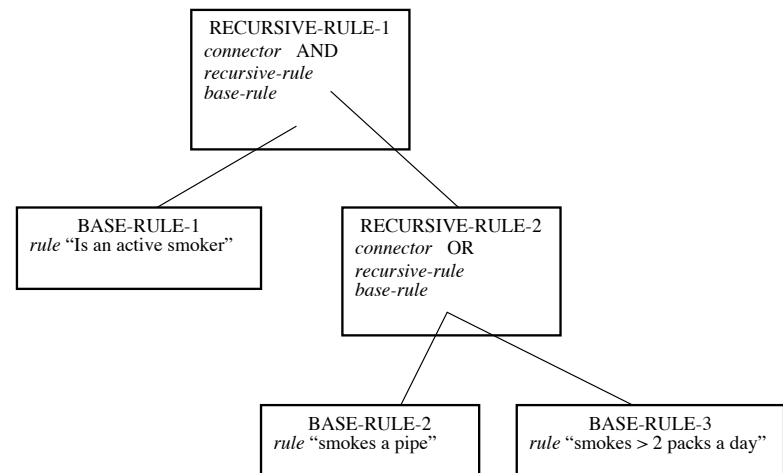
1. EB Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. Evidence-Based Medicine Working Group. JAMA 1992;268(17):2420-5.
2. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century: National Academy Press; 2001.
3. Berlin A, Sorani M, Sim I. Characteristics of Outpatient Clinical Decision Support Systems: A Taxonomic Description. under review 2004.
4. Friedman C, Hripscak G. Natural language processing and its future in medicine. Acad Med 1999;74(8):890-5.
5. The Trial Bank Project. Available at: <http://rctbank.ucsf.edu/>. Accessed February 13, 2003.
6. Sim I, Owens DK, Lavori PW, Rennels GD. Electronic Trial Banks: A Complementary Method for Reporting Randomized Trials. Med Decis Making 2000;20(4):440-450.
7. Sim I, Gorman P, Greenes R, Haynes R, Kaplan B, Lehmann H, et al. Clinical Decision Support Systems for the Practice of Evidence-Based Medicine. J Am Med Inform Assoc 2001;8(6):527-534.
8. Fernandez Lopez M. Overview of Methodologies for Building Ontologies. In: Benjamins VR, editor. IJCAL-99 Workshop on Ontologies and Problem-Solving Methods; 1999 August 2, 1999; Stockholm, Sweden: CEUR Publications; 1999.
9. Sim I, Cummings SR. A new framework for describing and quantifying the gap between proof and practice. Med Care 2003;41(8):874-81.
10. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999;282(15):1458-65.
11. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126(5):376-80.
12. Sim I, Sanders GD, McDonald KM. Evidence-based practice for mere mortals: the role of informatics and health services research. J Gen Intern Med 2002;17(4):302-8.
13. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med 2001;134(8):657-62.
14. Clarke MJ, Stewart LA. Obtaining data from randomised controlled trials: how much do we need for reliable and informative meta-analyses? BMJ 1994;309(6960):1007-10.
15. Meinert CL. Beyond CONSORT: need for improved reporting standards for clinical trials. Consolidated Standards of Reporting Trials. JAMA 1998;279(18):1487-9.
16. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med 1987;317(7):426-32.
17. Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. Controlled Clin Trials 1984;5(4):328-47.
18. CONSORT. Available at: <http://www.consort-statement.org/>. Accessed February 13, 2003.
19. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134(8):663-94.
20. Collaboration C. The Cochrane Library [CDROM]: Update Software; 2002.
21. ClinicalTrials.gov. National Institutes of Health. Available at: [www.clinicaltrials.gov](http://www.clinicaltrials.gov). Accessed September 16, 2002.
22. Current Controlled Trials -- Search the metaRegister of Controlled Trials. Current Controlled Trials. Available at: <http://controlled-trials.com/mrct/>. Accessed December 30, 2003.
23. Cancer.gov -- Search Clinical Trials. National Cancer Institute. Available at: [http://www.nci.nih.gov/search/clinical\\_trials/](http://www.nci.nih.gov/search/clinical_trials/). Accessed December 10, 2003.
24. Easterbrook PJ. Directory of registries of clinical trials. Stat Med 1992;11(3):345-59.
25. Tu S, Eriksson H, Gennari J, Shahar Y, Musen M. Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools: Application of Protege-II to protocol-based decision support. Technical Report: Knowledge Systems Lab; 1994 December 1994. Report No.: KSL-94-22.
26. Clinical Trials Object Model. National Cancer Institute. Available at: <http://ncicb.nci.nih.gov/content/coreftp/trials/Spore/Spore.html>. Accessed December 5, 2003.
27. Wyatt JC, Altman DG, Heathfield HA, Pantin CF. Development of Design-a-Trial, a knowledge-based critiquing system for authors of clinical trial protocols. Comput Methods Programs Biomed 1994;43(3-4):283-91.

28. Tu SW, Kemper CA, Lane NM, Carlson RW, Musen MA. A methodology for determining patients' eligibility for clinical trials. *Methods Inf Med* 1993;32(4):317-25.
29. Chandrasekaran B, Johnson T. Generic tasks and task structures: History, critique, and new directions. In: David J-M, Krivine J-P, Simmons R, editors. *Second-Generation Expert Systems*: Springer-Verlag; 1993. p. 232-272.
30. Gruninger M, Fox M. Methodology for the design and evaluation of ontologies. Dept. of Industrial Engineering Technical Report: University of Toronto; 1995 April 13.
31. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol* 1995;48(1):167-71.
32. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't [editorial] [see comments]. *BMJ* 1996;312(7023):71-2.
33. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. *Methods Inf Med* 1993;32(4):281-91.
34. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991;44(11):1271-8.
35. Gibaldi M, Sullivan S. Intention-to-treat analysis in randomized trials: who gets counted? *J Clin Pharmacol* 1997;37(8):667-72.
36. Gottlieb SS. Dead is dead--artificial definitions are no substitute [comment] [see comments]. *Lancet* 1997;349(9053):662-3.
37. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16(1):62-73.
38. Karp P, Chaudhri V, Paley S. A Collaborative Environment for Authoring Large Knowledge Bases. *Journal of Intelligent Information Systems* 1998:In press.
39. Uschold M, Gruninger M. Ontologies: principles, methods and applications. *Knowledge Engineering Review* 1996;11(2):93-136.
40. NCI Common Data Elements Dictionary. National Cancer Institute. Available at: [http://ccl-server5.nci.nih.gov:8080/pls/cde\\_public/cde\\_java.show](http://ccl-server5.nci.nih.gov:8080/pls/cde_public/cde_java.show). Accessed October 3, 2002.
41. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. *J Am Med Inform Assoc* 1996;3(3):224-33.

42. Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. *J Am Med Inform Assoc* 1997;4(6):484-500.
43. Goldberg H, Goldsmith D, Law V, Keck K, Tuttle M, Safran C. An evaluation of UMLS as a controlled terminology for the Problem List Toolkit. *Medinfo* 1998;9 Pt 1:609-12.
44. Tuttle MS, Olson NE, Keck KD, Cole WG, Erlbaum MS, Sherertz DD, et al. Metaphrase: an aid to the clinical conceptualization and formalization of patient problems in healthcare enterprises. *Methods Inf Med* 1998;37(4-5):373-83.
45. Elkin PL, Tuttle M, Keck K, Campbell K, Atkin G, Chute CG. The role of compositionality in standardized problem list generation. *Medinfo* 1998;9 Pt 1:660-4.
46. Musen MA, Tu SW, Das AK, Shahar Y. EON: a component-based approach to automation of protocol-directed therapy. *J Am Med Inform Assoc* 1996;3(6):367-88.
47. Re-engineering the Clinical Research Enterprise. National Institutes of Health. Available at: <http://nihroadmap.nih.gov/clinicalresearch/index.asp>. Accessed December 16, 2003.
48. Health Level 7: Regulated Clinical Research Information Management. Health Level 7. Available at: <http://www.hl7.org/Special/committees/rcrim/rcrim.htm>. Accessed October 3, 2002.



**Figure 1. Lifecycle View of Clinical Trials for Evidence-Based Practice.** This figure illustrates the place of clinical trials in the evidence chain for evidence-based practice. After trials are performed, their results must be interpreted and applied to clinical care. For the scientific community to properly interpret trials, both the trial protocols and results should be available, for example in repositories such as ClinicalTrials.gov and RCT Bank respectively. Results from related trials should then be synthesized into systematic reviews (e.g., as collected in the Cochrane Collaboration database), incorporated into decision models and guidelines (e.g., as collected in the National Guideline Clearinghouse), and finally matched to electronic medical records to facilitate evidence application.



**Figure 2. Representation of an Entrance Rule.** This rule representation captures the entrance criterion “Is an active smoker AND (smokes a pipe OR smokes more than 2 packs/day).”

Recursive rules allow composition and nesting of base rules. Instance names are all uppercase; slot names are in italics; lines point to slot values.



**Table 1. Upper Level Subtasks of Systematic Review Task Hierarchy**

First-Level Subtask	Second-Level Subtasks
Trial Retrieval	A. Capture the query B. Match the query to the trials
Judging Internal Validity	A. Was the statistical design of the trial appropriate? B. Was there any intervention assignment bias? C. Were the intervention groups comparable? D. Was there any intervention-related bias? E. Were there co-interventions that may have confounded the results? F. Are the outcome variables meaningful? G. Was there any outcome assessment or measurement bias? H. Was there any follow-up bias? I. Were the results analyzed appropriately? J. What biases might the trial personnel have introduced? K. Is the trial internally valid?
Judging External Validity	A. Is the study of interest to the target population? B. Are the trial subjects and the target population likely to belong to the same overall population? C. Is the setting of the trial comparable to the setting of the target population's intervention?

First-Level Subtask	Second-Level Subtasks
	D. Is the tested intervention reproducible for the target population?
Quantitative Synthesis	A. Is it appropriate to combine these quantitative results? B. Calculate summary statistic for pair-wise comparisons C. Quantitative meta-analysis
Contextual Interpretation	A. What is the scientific and clinical context of the study? B. What is the scientific discourse on the study question? C. What is the social, economic, ethical, and legal context of this study?

**Table 2. Examples of Trial Information Modeled in RCT Schema**

Concept Group	Examples
Administration	Trial title, stage, dates, and investigators, study sites, funding, ethics, description of trial committees, errata
Background	Text description of study background, objectives, rationale
Design	Design of trial, statistics used, details of randomization and allocation concealment, subgroup information
Entrance Criteria	Inclusion and exclusion criteria
Interventions	Description of interventions and co-interventions, details of treatment masking and administration
Outcome Variables	Definitions of outcome variables, baseline characteristics, side effect variables, details of outcome analysis, assessment, and measurement
Enrollment	Recruitment, screening, enrollment of subjects, number of subjects in each intervention group
Follow-up and Compliance	Follow-up of subjects, compliance, crossovers
Results	Quantitative study results
Conclusions and Publications	Discussion of study's limitations and conclusions, citations to trial publications

**Table 3. Characteristics of RCT Bank Trials**

Characteristic	Examples
Clinical Domains	Cardiology, radiology, geriatrics, psychiatry
Intervention Types	Procedures (e.g., thrombolysis), single and multi-step drugs (e.g., aspirin, warfarin), counseling, multiple interventions in one arm (e.g., 8 fall prevention interventions)
Outcome Types	Dichotomous, continuous, univariate, multivariate, survival, regression, scored instruments (e.g., Wechsler Memory Scale)
Result Types	Intention-to-treat and efficacy analyses, subgroup analyses